MTH 8302 - Modèles de Régression et d'Analyse de Variance

Leçon 4 : Modèles Linéaires Généralisés et Méthodes Classiques d'Apprentissage Supervisé

Polytechnique Montréal - Hiver 2025

Chiheb Trabelsi

April 2, 2025

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Table des Matières

Modèles Linéaires Généralisés

Modèles Linéaires Généralisés : Introduction

Modèles Linéaires Généralisés : Introduction

- Dans le cadre de la régression linéaire, l'objectif est de modéliser une relation entre :
  - une variable réponse continue $Y \in \mathbb{R} \Rightarrow \mathbf{Y} \in \mathbb{R}^n$
  - à partir de variables explicatives $X_1, \ldots, X_p \in \mathbb{R}^n \Rightarrow \mathbf{X} \in \mathbb{R}^{n\times(p+1)}$.
- Cette relation est exprimée à l'aide d'un modèle de linéaire :

$$Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{où } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n),$$

  - où $\boldsymbol{\beta} \in \mathbb{R}^{(p+1)}$ est le vecteur des paramètres à estimer.
- L'estimation des paramètres se fait typiquement par moindres carrés ou maximum de vraisemblance, ce qui revient à trouver l'estimateur $\hat{\boldsymbol{\beta}}$, tel que:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

- Cette approche est efficace **si les hypothèses sont vérifiées :**
  - Simple à mettre en oeuvre.
  - Interprétable puisque chaque β_j mesure l'effet linéaire de X_j.

Modèles Linéaires Généralisés : Introduction

- Cependant, les hypothèses considérées dans cette approche sont fortes :
  - $Y \in \mathbb{R}$ est nécessairement continue.
  - Les erreurs ϵ sont gaussiennes, centrées et homoscédastiques.
  - La relation entre Y et les X_j est linéaire.
- **Que se passe-t-il si Y n'est pas continue ?**
- **Exemples réalistes :**
  - $Y \in \{0, 1\}$ (binaire) : Prédire si un patient est malade (oui/non), si un client remboursera son prêt, etc.
  - $Y \in \mathbb{N}$: Nombre d'appels reçus dans un call center, nombre d'accidents dans une intersection, etc.
- **Problèmes de la régression linéaire dans ces cas :**
  - **Incohérence** : Prédictions négatives ou > 1.
  - **Statistiquement sous-optimale** : Ne tient pas compte de la distribution naturelle de Y.
  - **Peu performante** : Mauvaise généralisation.

Modèles Linéaires Généralisés : Introduction

- L'objectif est de généraliser les modèles linéaires classiques pour les rendre compatibles avec des situations où :
  - La variable cible Y n'est pas continue.
  - On veut modéliser des probabilités, des décomptes, voire même des proportions.
  - On souhaite garder une structure linéaire dans l'espace des variables explicatives X_j, tout en respectant la distribution naturelle de Y.
- **Solution :** Introduire les **Modèles Linéaires Généralisés** (GLM : *Generalized Linear Models*), qui permettent de :
  - Étendre la régression linéaire à **d'autres types de variables cibles.**
  - Garder une structure linéaire dans l'espace des prédicteurs X_j. Cela repose sur une idée simple mais efficace : Relier $\mathbb{E}[Y \mid X_1, \ldots, X_p]$, l'espérance de Y à une combinaison linéaire des variables explicatives X_j **via une fonction de lien adaptée notée g.**
  - **Tenir compte de la distribution de Y.** Cela permet d'intégrer des modèles comme :
    - La régression logistique (Y binaire $\Rightarrow Y \sim$ Bernoulli(p)).
    - La régression de Poisson (décompte $\Rightarrow Y \sim$ Poisson(λ)).
    - Etc.

La Famille Exponenielle des Distributions et Composantes des Modèles Linéaires Généralisés

La Famille Exponenielle et Composantes des GLM

- Notation adoptée :
  - $\mathbf{Y} \in \mathbb{R}^n$ et $Y \in \mathbb{R}$: variables aléatoires.
  - $\boldsymbol{y} \in \mathbb{R}^n$ et $y \in \mathbb{R}$: Réalisations (valeurs observées) respectives des variables aléatoire $\mathbf{Y}$ et Y

  Dans une fonction de densité, on écrit :

$$f_Y(y; \theta, \phi)$$

  - $f_Y(\cdot)$ désigne la densité de la variable aléatoire Y,
  - y est la valeur sur laquelle la densité est évaluée.

- L'expression de la densité de la famille exponentielle est donnée par :

$$f_Y(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right).$$

  - Y est bien la variable aléatoire (indiquée dans l'indice de f_Y),
  - y est la variable d'intégration ou d'évaluation (la réalisation de Y).

La Famille Exponenielle et Composantes des GLM

- L'expression de la densité de la famille exponentielle est donnée par :

$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right).$$

  Signification des différentes composantes :
  - y : réalisation de la variable aléatoire Y
  - θ : paramètre naturel (ou canonique) de la distribution
  - ϕ : paramètre de dispersion (ou fixé à 1 selon le contexte)
  - $a(\phi)$: fonction de pondération de la dispersion (souvent $a(\phi) = \phi$)
  - $b(\theta)$: fonction log-partition (fonction de normalisation logarithmique) qui garantit l'intégrabilité à 1. Elle détermine la moyenne et la variance de Y.
  - $c(y,\phi)$: fonction indépendante de θ, absorbant les termes restants pour assurer la normalisation.
- **Propriétés des Moments de la distribution (Preuve Ici):**
  - $\mathbb{E}[Y] = \frac{\partial b(\theta)}{\partial \theta} = b'(\theta)$
  - $\text{Var}(Y) = a(\phi)\frac{\partial b(\theta)}{\partial^2 \theta} = a(\phi) \cdot b''(\theta)$
- **Remarque :** Cette forme permet d'exprimer une large famille de distributions usuelles dans un même cadre probabiliste.

La Famille Exponenielle et Composantes des GLM

Un GLM est défini par 3 composantes fondamentales :

1. **La distribution de la variable cible Y**
   - Y suit une distribution **appartenant à la famille exponentielle :**
     - **loi normale** (régression linéaire)
     - **loi de Bernoulli** (régression logistique)
     - **loi de Poisson** (décomptes)
     - autres : binomiale, gamma, inverse-gaussienne, etc.
   - Forme générale :

$$f_Y(y;\theta) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

2. **La fonction de lien g**
   - Relie $\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y} \mid \mathbf{X}]$ au prédicteur linéaire $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$
   - $g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \quad$ (avec g appliqué élément par élément).
   - Le choix du lien dépend de la distribution de Y :
     - **Logit** pour Bernoulli
     - **Log** pour Poisson
     - **Identité** pour normale
3. **Le prédicteur linéaire η**
   - $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$.
   - Permet une estimation par maximum de vraisemblance.

La Famille Exponenielle et Composantes des GLM

2. La fonction de lien g

- Pour une $i$ème observation individuelle $\mathbf{x}_i$, la fonction de lien relie l'espérance conditionnelle $\mu_i = \mathbb{E}[Y_i \mid \mathbf{X}_i = \mathbf{x}_i]$ au prédicteur linéaire $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$:

$$g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$$

- Pour l'ensemble des n observations, on définit les vecteurs :

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y} \mid \mathbf{X}], \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

- La fonction de lien s'applique alors élément par élément :

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta}$$

3. Le prédicteur linéaire $\boldsymbol{\eta}$

- Pour une observation : $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$
- Pour l'échantillon : $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$
- Il constitue le coeur du modèle linéaire généralisé et permet de relier l'espace des variables explicatives à $\boldsymbol{\mu} = \mathbb{E}[\mathbf{Y} \mid \mathbf{X}]$.
- Il permet l'estimation de $\boldsymbol{\beta}$ par maximum de vraisemblance.

Étapes pour Vérifier qu'une loi appartient à la famille exp

- **Objectif :** Vérifier si la densité $f_Y(y)$ appartient à la famille exponentielle des distributions et pourrait donc s'écrire comme:

$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

- **Étapes à suivre :**
  1. **Écrire la densité dans sa forme classique.**
     Exemple : loi normale, Bernoulli, Poisson...
  2. **Mettre la densité sous forme exponentielle.**
     Identifier les termes permettant d'aboutir à la forme canonique.
  3. **Identifier les composantes du modèle exponentiel :**
     - θ : paramètre naturel
     - ϕ : paramètre de dispersion
     - $a(\phi)$: fonction de pondération
     - $b(\theta)$: fonction log-partition
     - $c(y,\phi)$: fonction de normalisation

Étapes pour Vérifier qu'une loi appartient à la famille exp

4. **Calculer les moments :**

$$\mathbb{E}[Y] = b'(\theta), \quad \text{Var}(Y) = a(\phi) \cdot b''(\theta)$$

5. **Déterminer la fonction de lien canonique :**

$$g(\mu) = \theta \quad \Rightarrow \quad \mu = b'(\theta)$$

6. **Exprimer le prédicteur linéaire du GLM :**

$$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad \text{avec } g(\mu_i) = \eta_i$$

7. **Conclure :**
   - La loi appartient à la famille exponentielle.
   - Elle peut être modélisée via un GLM avec :
     - Fonction de lien : $g(\cdot)$

Exercice : Montrons que la loi normale de $Y \sim \mathcal{N}(\mu, \sigma^2)$ appartient à la famille exponentielle des densités.

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q1. Quelle est la densité d'une variable $Y \sim \mathcal{N}(\mu, \sigma^2)$?

A. $\exp\left(-\frac{(y-\mu)^2}{\sigma^2}\right)$

B. $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$

C. $\frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q1. Quelle est la densité d'une variable $Y \sim \mathcal{N}(\mu, \sigma^2)$?
Réponse correcte : B
La densité d'une loi normale est bien :

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q2. Développer l'expression $(y-\mu)^2$

A. $y^2+\mu^2$

B. $y^2-2y\mu+\mu^2$

C. $y^2+2y\mu+\mu^2$

D. $y^2-2\mu y+\mu^2$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q2. Développer l'expression $(y-\mu)^2$
Réponse correcte : D
On a : $(y-\mu)^2 = y^2 - 2y\mu + \mu^2$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q3. Mise sous forme exponentielle : quelle est la bonne forme ?

A. $\exp\left(-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \log(2\pi\sigma^2)\right)$

B. $\exp\left(\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)$

C. $\exp\left(\frac{(y-\mu)^2}{2\sigma^2}\right)$

D. $\exp\left(\frac{(y-\mu)^2}{\sigma^2}\right)$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q3. Mise sous forme exponentielle : quelle est la bonne forme ?

Réponse correcte : B

C'est la bonne mise sous la forme : $\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q4. Quelles sont les expressions correctes ?

A. $\theta = \mu$, $a(\phi) = \phi = \sigma^2$, $b(\theta) = \frac{1}{2}\theta^2$

B. $\theta = y$, $b(\theta) = \mu^2$, $a(\phi) = \sigma^2$

C. $\theta = \sigma^2$, $b(\theta) = \theta^2$, $a(\phi) = \mu$

D. $\theta = \mu$, $b(\theta) = \mu^2$, $a(\phi) = \phi^2$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q4. Quelles sont les expressions correctes ?

Réponse correcte : A

On identifie bien : $\theta = \mu$, $b(\theta) = \frac{1}{2}\theta^2$, $a(\phi) = \phi = \sigma^2$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q5. Que valent les moments de Y ?

A. $\mathbb{E}[Y] = \theta$, $\text{Var}(Y) = \phi$

B. $\mathbb{E}[Y] = b''(\theta)$, $\text{Var}(Y) = \phi b'(\theta)$

C. $\mathbb{E}[Y] = b'(\theta) = \theta$, $\text{Var}(Y) = a(\phi)b''(\theta) = \phi$

D. $\mathbb{E}[Y] = \mu$, $\text{Var}(Y) = \mu^2$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q5. Que valent les moments de Y ?

Réponses correctes : A et C

$\mathbb{E}[Y] = b'(\theta) = \theta = \mu$, $\text{Var}(Y) = a(\phi) \cdot b''(\theta) = \phi$

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q6. La régression linéaire est-elle un GLM ?

A. GLM avec loi normale, lien log

B. GLM avec loi normale, lien identité

C. Ce n'est pas un GLM car Y est continu

D. GLM avec fonction de lien logit

Exercice : La Loi Normale $\in$ Famille Exponentielle ?

Q6. La régression linéaire est-elle un GLM ?
Réponse correcte : B
Régression linéaire = GLM avec loi normale et la fonction identité comme lien.

Solution de l'Exercice : Montrons que la loi normale de $Y \sim \mathcal{N}(\mu, \sigma^2)$ appartient à la famille exponentielle des densités.

Solution : La Loi Normale $\in$ Famille Exponentielle ?

- Montrons que la loi normale de $Y \sim \mathcal{N}(\mu, \sigma^2)$ **appartient à la famille exponentielle des densités:**

$$f_Y(y;\theta) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

- **Étape 1 : Écrire la densité de la loi normale**

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

- **Étape 2 : Mise sous forme exponentielle**

$$\begin{aligned}
f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2 - 2y\mu + \mu^2}{2\sigma^2}\right) \\
&= \exp\left(-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right) \\
&= \exp\left(\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)
\end{aligned}$$

Solution : La Loi Normale $\in$ Famille Exponentielle ?

- **Étape 2 : Mise sous forme exponentielle**

$$
\begin{aligned}
f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2 - 2y\mu + \mu^2}{2\sigma^2}\right) \\
&= \exp\left(-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right) \\
&= \exp\left(\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right) \\
&= \exp\left(\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)
\end{aligned}
$$

On reconnaît :

$$
f_Y(y) = \exp\left(\underbrace{\frac{y\mu - \mu^2/2}{\sigma^2}}_{\frac{y\theta - b(\theta)}{a(\phi)}} + \underbrace{-\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)}_{c(y,\phi)}\right)
$$

Solution : La Loi Normale $\in$ Famille Exponentielle ?

- **Étape 3 : Identification des termes** On reconnaît :

$$f_Y(y) = \exp\left(\underbrace{\frac{y\mu - \mu^2/2}{\sigma^2}}_{\frac{y\theta - b(\theta)}{a(\phi)}} + \underbrace{-\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)}_{c(y,\phi)}\right)$$

- **Identification :**
  - $\theta = \mu$: le paramètre canonique est l'espérance μ.
  - $a(\phi) = \phi = \sigma^2$, $b(\theta) = \frac{1}{2}\theta^2$.
  - $c(y,\phi) = -\frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi)$.
  - Fonction de lien canonique : $g(\mu) = \theta = \mu \Rightarrow g(\mu) = \mu$.
- **Donc :**
  - Le lien est l'identité : $\mu = \eta$
  - Le prédicteur linéaire est $\eta = \mathbf{x}_i^\top \boldsymbol{\beta} \Rightarrow \mu = \mathbf{x}_i^\top \boldsymbol{\beta}$

Conclusion : La régression linéaire est un GLM avec :

$$Y_i \sim \mathcal{N}(\mu_i, \sigma^2), \quad \mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$$

Solution : La Loi Normale $\in$ Famille Exponentielle ?

Composante du GLM	**Régression linéaire**
Distribution de Y	$Y \sim \mathcal{N}(\mu, \sigma^2)$
Paramètre naturel θ	$\theta = \mu$
Paramètre de dispersion ϕ	$\phi = \sigma^2$
Fonction de lien $g(\mu)$	$g(\mu) = \mu$ (identité)
Prédicteur linéaire η	$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$
Lien entre μ_i et η_i	$\mu_i = \eta_i \Rightarrow \mu = \mathbf{x}_i^\top \boldsymbol{\beta}$
Méthode d'estimation	Maximum de vraisemblance (OLS)

Conclusion : La régression linéaire *est* un GLM avec loi normale, lien identité, estimation par moindres carrés.

Exercice : Montrons que la loi de Poisson $Y \sim \mathcal{P}(\lambda)$ appartient à la famille exponentielle.

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q1. Quelle est la fonction de masse de $Y \sim \mathcal{P}(\lambda)$?

A. $f_Y(y) = \lambda^y \exp(-\lambda)$

B. $f_Y(y) = \dfrac{\lambda^y}{y!} \exp(-\lambda)$

C. $f_Y(y) = \exp(-\lambda y)$

D. $f_Y(y) = \dfrac{1}{y!} \exp(\lambda^y - \lambda)$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q1. Quelle est la fonction de masse de $Y \sim \mathcal{P}(\lambda)$?
Réponse correcte : B
La fonction de masse est :

$$f_Y(y) = \frac{\lambda^y}{y!} \exp(-\lambda), \quad y \in \mathbb{N}$$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q2. Mise sous forme exponentielle : quelle est la bonne forme ?

A. $\exp\left(y \log \lambda - \lambda - \log y!\right)$

B. $\exp\left(\lambda y - \lambda^2 - \log y!\right)$

C. $\exp\left(y\lambda - \lambda - y^2\right)$

D. $\exp\left(y + \lambda - y \log y\right)$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q2. Mise sous forme exponentielle : quelle est la bonne forme ?
Réponse correcte : A
On reconnaît la forme exponentielle :

$$f_Y(y) = \exp\left(y \log \lambda - \lambda - \log y!\right)$$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q3. Quels sont les bons choix pour les composantes du modèle exponentiel ?

A. $\theta = \log \lambda$, $b(\theta) = e^{\theta}$, $a(\phi) = 1$

B. $\theta = \lambda$, $b(\theta) = \log \theta$, $a(\phi) = \phi = \lambda$

C. $\theta = \lambda$, $b(\theta) = \theta$, $a(\phi) = 1$

D. $\theta = \log \lambda$, $b(\theta) = \theta^2$, $a(\phi) = \lambda$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q3. Quels sont les bons choix pour les composantes du modèle exponentiel ?

Réponse correcte : A

En effet, on a :

$$\boxed{\theta = \log \lambda, \quad b(\theta) = e^{\theta} = \lambda, \quad a(\phi) = 1}$$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q4. Que valent les moments de $Y \sim \mathcal{P}(\lambda)$?

A. $\mathbb{E}[Y] = \theta$, $\mathsf{Var}(Y) = \theta$

B. $\mathbb{E}[Y] = b'(\theta) = e^{\theta}$, $\mathsf{Var}(Y) = b''(\theta) = e^{\theta}$

C. $\mathbb{E}[Y] = b''(\theta)$, $\mathsf{Var}(Y) = a(\phi)b'(\theta)$

D. $\mathbb{E}[Y] = \log(\lambda)$, $\mathsf{Var}(Y) = \lambda^2$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q4. Que valent les moments de $Y \sim \mathcal{P}(\lambda)$?
Réponse correcte : B
On a :

$$\mathbb{E}[Y] = b'(\theta) = e^{\theta} = \lambda, \quad \mathrm{Var}(Y) = b''(\theta) = e^{\theta} = \lambda$$

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q5. La régression de Poisson est-elle un GLM ?

A. GLM avec loi de Poisson, lien identité

B. GLM avec loi de Poisson, lien carré

C. GLM avec loi de Poisson, lien log

D. Ce n'est pas un GLM car Y est discrète

Exercice : La Loi de Poisson $\in$ Famille Exponentielle ?

Q5. La régression de Poisson est-elle un GLM ?
Réponse correcte : C
Lien canonique de la loi de Poisson : $g(\mu) = \log \mu$ (car $\theta = \log \lambda = \log \mu$)

Solution de l'Exercice : Montrons que la loi de Poisson $Y \sim \mathcal{P}(\lambda)$ appartient à la famille exponentielle.

Solution : La Loi de Poisson $\in$ Famille Exponentielle ?

- **Étape 1 : Écrire la densité de la loi de Poisson $Y \sim \mathcal{P}(\lambda)$:**

$$f_Y(y) = \frac{\lambda^y}{y!} e^{-\lambda}, \quad y \in \mathbb{N}$$

- **Étape 2 : Mettre l'expression sous forme exponentielle :**

$$f_Y(y) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)$$

- On écrit alors:

$$f_Y(y) = \frac{\lambda^y}{y!} e^{-\lambda} = \exp\left(y \log \lambda - \lambda - \log y!\right)$$

- **Étape 3 : Identification des termes**

  Cela ressemble à :

$$f_Y(y) = \exp\left(y\theta - b(\theta) + c(y)\right),$$

  avec :

$$\theta = \log \lambda, \quad b(\theta) = e^{\theta} = \lambda, \quad c(y) = -\log y!$$

Solution : Moments

- On vient d'exprimer $f_Y(y)$ sous forme exponentielle où l'on a:

$$f_Y(y) = \exp\left(y\theta - b(\theta) + c(y)\right)$$

avec :

$$\theta = \log\lambda, \quad b(\theta) = e^{\theta} = \lambda, \quad c(y) = -\log y!$$

- Dans la forme canonique exponentielle, on a :

$$\mathbb{E}[Y] = b'(\theta), \quad \text{Var}(Y) = a(\phi) \cdot b''(\theta)$$

- Or :

$$b(\theta) = e^{\theta} \Rightarrow b'(\theta) = e^{\theta} = \lambda, \quad b''(\theta) = e^{\theta} = \lambda$$

- Donc :

$$\mathbb{E}[Y] = \lambda, \quad \text{Var}(Y) = 1 \cdot \lambda = \lambda$$

Solution : La Loi de Poisson $\in$ Famille Exponentielle ?

- **Étape 3 : Identification des termes**
  On reconnaît :

$$f_Y(y) = \exp\left(\underbrace{y\log\lambda - \lambda}_{y\theta - b(\theta)} + \underbrace{-\log y!}_{c(y)}\right)$$

- **Identification :**
  - $\theta = \log\lambda$: le paramètre canonique est le logarithme de l'espérance.
  - $b(\theta) = e^{\theta} = \lambda$
  - $a(\phi) = 1$, car il n'y a pas de paramètre de dispersion dans le modèle de Poisson.
  - $c(y, \phi) = -\log y!$
  - $\mathbb{E}[Y] = \lambda = \mu$.
  - Fonction de lien canonique :
    $g(\mu) = \theta = \log\lambda = \log\mu \Rightarrow g(\mu) = \log\mu$
- **Donc :**
  - Le lien est logarithmique : $\log\mu = \eta$
  - Le prédicteur linéaire est : $\eta = \mathbf{x}_i^\top\boldsymbol{\beta} \Rightarrow \mu = \exp(\mathbf{x}_i^\top\boldsymbol{\beta})$.

Récapitulatif : Poisson = GLM

Composante du GLM	**Régression de Poisson**
Distribution de Y	$Y \sim \mathcal{P}(\mu)$
Paramètre naturel θ	$\theta = \log \mu$
Paramètre de dispersion ϕ	$\phi = 1$
Fonction de lien $g(\mu)$	$g(\mu) = \log \mu$
Prédicteur linéaire η	$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$
Lien entre μ_i et η_i	$\mu_i = \exp(\eta_i) \Rightarrow \mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$
Méthode d'estimation	Maximum de vraisemblance

Conclusion : La régression de Poisson est un GLM avec lien log et distribution de Poisson :

$$Y_i \sim \mathcal{P}(\mu_i), \quad \mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}).$$

Exercice : Estimation par Maximum de Vraisemblance pour la Régression de Poisson

QCM 1 : MV pour la Régression de Poisson

Question : Quelle est l'expression correcte de la fonction de masse de probabilité pour une variable aléatoire $Y_i \sim \mathcal{P}(\mu_i)$?

A. $\mathbb{P}(Y_i = y_i) = \mu_i^{y_i} \exp(-\mu_i)$

B. $\mathbb{P}(Y_i = y_i) = \frac{\mu_i^{y_i} \exp(-\mu_i)}{y_i!}$

C. $\mathbb{P}(Y_i = y_i) = \frac{y_i^{\mu_i} \exp(-y_i)}{\mu_i!}$

D. $\mathbb{P}(Y_i = y_i) = \mu_i \cdot \exp(-y_i)$

QCM 1 : MV pour la Régression de Poisson

Question : Quelle est l'expression correcte de la fonction de masse de probabilité pour une variable aléatoire $Y_i \sim \mathcal{P}(\mu_i)$?
Bonne réponse : B

Justification : C'est la définition exacte de la loi de Poisson :

$$\mathbb{P}(Y_i = y_i) = \frac{\mu_i^{y_i} \exp(-\mu_i)}{y_i!}$$

QCM 2 : MV pour la Régression de Poisson

Question : Dans le modèle de régression de Poisson, quel lien exprime correctement l'espérance μ_i en fonction des variables explicatives $\mathbf{x}_i$?

A. $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$

B. $\mu_i = \log(\mathbf{x}_i^\top \boldsymbol{\beta})$

C. $\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$

D. $\mu_i = \sqrt{\mathbf{x}_i^\top \boldsymbol{\beta}}$

QCM 2 : MV pour la Régression de Poisson

Question : Dans le modèle de régression de Poisson, quel lien exprime correctement l'espérance μ_i en fonction des variables explicatives $\mathbf{x}_i$?
Bonne réponse : C

Justification : C'est la fonction de lien canonique utilisée pour les modèles de régression de Poisson.

QCM 3 : MV pour la Régression de Poisson

Question : Quelle est la forme correcte de la log-vraisemblance dans le modèle de régression de Poisson (à un facteur constant près) ?

A. $\sum_{i=1}^{n} [y_i \eta_i - \exp(\eta_i)]$
B. $\sum_{i=1}^{n} [y_i \cdot \log(\eta_i) - \eta_i]$
C. $\sum_{i=1}^{n} [y_i \cdot \exp(\eta_i) - \eta_i]$
D. $\sum_{i=1}^{n} \left[y_i^2 - \eta_i\right]$

QCM 3 : MV pour la Régression de Poisson

Question : Quelle est la forme correcte de la log-vraisemblance dans le modèle de régression de Poisson (à un facteur constant près) ?
Bonne réponse : A

Justification : La log-vraisemblance est donnée (à un terme constant près) par :

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} [y_i \eta_i - \exp(\eta_i)]$$

QCM 4a : MV pour la Régression de Poisson

Question : Trouvez $\frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{x}_i^\top \boldsymbol{\beta}$.

A. $\boldsymbol{\beta}$

B. $\mathbf{x}_i$

C. $\mathbf{x}_i^\top$

D. 1

QCM 4a : MV pour la Régression de Poisson

Question : Trouvez $\frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{x}_i^\top \boldsymbol{\beta}$.
Bonne réponse : B

Justification : La dérivée de $\mathbf{x}_i^\top \boldsymbol{\beta}$ par rapport à $\boldsymbol{\beta}$ est le vecteur $\mathbf{x}_i$, car c'est une fonction linéaire.

QCM 4b : MV pour la Régression de Poisson

Question : Trouvez $\frac{\partial}{\partial \boldsymbol{\beta}} \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$.

A. $\mathbf{x}_i$

B. $\exp(\mathbf{x}_i^\top \boldsymbol{\beta})$

C. $\exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \cdot \mathbf{x}_i$

D. $\mathbf{x}_i^\top \cdot \exp(\boldsymbol{\beta})$

QCM 4b : MV pour la Régression de Poisson

Question : Trouvez $\frac{\partial}{\partial \boldsymbol{\beta}} \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$.
Bonne réponse : C

Justification : C'est une application de la règle de la chaîne :

$$\frac{\partial}{\partial \boldsymbol{\beta}} \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \cdot \mathbf{x}_i$$

QCM 4c : MV pour la Régression de Poisson

Question : Quelle est l'expression correcte du gradient de la log-vraisemblance par rapport à $\boldsymbol{\beta}$, $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$?

A. $\sum_{i=1}^{n} y_i \cdot \mathbf{x}_i$

B. $\sum_{i=1}^{n} \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \cdot \mathbf{x}_i$

C. $\sum_{i=1}^{n} (y_i - \exp(\mathbf{x}_i^\top \boldsymbol{\beta})) \cdot \mathbf{x}_i$

D. $\sum_{i=1}^{n} (y_i + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})) \cdot \mathbf{x}_i$

QCM 4c : MV pour la Régression de Poisson

Question : Quelle est l'expression correcte du gradient de la log-vraisemblance par rapport à $\boldsymbol{\beta}$, $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$?
Bonne réponse : C

Justification : En combinant les dérivées des deux termes :

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left[y_i - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \right] \mathbf{x}_i$$

QCM 5 : MV pour la Régression de Poisson

Question : Pourquoi ne peut-on pas obtenir de solution fermée pour $\hat{\beta}$ dans le modèle de régression de Poisson ?

A. Car la log-vraisemblance dépend de $\log(y_i!)$

B. Car les dérivées impliquent des fonctions exponentielles

C. Car les variables $\mathbf{x}_i$ sont corrélées

D. Car y_i peut prendre des valeurs nulles

QCM 5 : MV pour la Régression de Poisson

Question : Pourquoi ne peut-on pas obtenir de solution fermée pour $\hat{\boldsymbol{\beta}}$ dans le modèle de régression de Poisson ?

Bonne réponse : B

Justification : Les termes exponentiels dans le gradient empêchent de résoudre explicitement le système $\nabla \ell(\boldsymbol{\beta}) = 0$.

Solution : Estimation par Maximum de Vraisemblance pour la Régression de Poisson

Solution : MV pour la Régression de Poisson

- **Loi de Poisson (définition)** : Soit une variable aléatoire $Y_i \sim \mathcal{P}(\mu_i)$, la fonction de masse de probabilité est :

$$\mathbb{P}(Y_i = y_i \mid \mu_i) = \frac{\mu_i^{y_i} \exp(-\mu_i)}{y_i!}, \quad y_i \in \mathbb{N}$$

- **Lien avec les variables explicatives : Régression de Poisson** Dans un modèle de régression de Poisson (GLM), on relie l'espérance μ_i aux variables explicatives via une fonction de lien canonique :

$$\log(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \Rightarrow \quad \mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$$

- **Substitution dans la densité :** On remplace alors μ_i dans la loi de Poisson par $\exp(\mathbf{x}_i^\top \boldsymbol{\beta})$, ce qui donne :

$$\mathbb{P}(Y_i = y_i \mid \mathbf{x}_i) = \frac{\left[\exp(\mathbf{x}_i^\top \boldsymbol{\beta})\right]^{y_i} \exp\left(-\exp(\mathbf{x}_i^\top \boldsymbol{\beta})\right)}{y_i!}$$

- **Utilité** : Cette expression est la base de la construction de la fonction de vraisemblance pour le modèle.

Solution : MV pour la Régression de Poisson

- On a donc :

$$\mathbb{P}(Y_i = y_i \mid \mathbf{x}_i) = \frac{[\exp(\mathbf{x}_i^\top \boldsymbol{\beta})]^{y_i} \exp(-\exp(\mathbf{x}_i^\top \boldsymbol{\beta}))}{y_i!}$$

- Hypothèse : observations indépendantes $\Rightarrow$ vraisemblance :

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{[\exp(\mathbf{x}_i^\top \boldsymbol{\beta})]^{y_i} \exp(-\exp(\mathbf{x}_i^\top \boldsymbol{\beta}))}{y_i!}$$

- On prend la fonction log-vraisemblance :

$$\ell(\boldsymbol{\beta}) = \log \mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \mathbf{x}_i^\top \boldsymbol{\beta} - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) - \log(y_i!) \right]$$

- La fonction objectif à maximiser est alors donnée par :

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \eta_i - \exp(\eta_i) - \log(y_i!) \right], \quad \text{où } \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$$

- Le terme $\log(y_i!)$ ne dépend pas de $\boldsymbol{\beta}$, on peut l'ignorer pour l'optimisation.

Solution : MV pour la Régression de Poisson

- On fonction log-vraisemblance pour des observations indépendantes est donnée par :

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \cdot \mathbf{x}_i^\top \boldsymbol{\beta} - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) - \log(y_i!) \right].$$

- **Calcul du gradient :** dérivons chaque terme par rapport à $\boldsymbol{\beta}$:

$$\frac{\partial}{\partial \boldsymbol{\beta}} \left(y_i \cdot \mathbf{x}_i^\top \boldsymbol{\beta} \right) = y_i \cdot \mathbf{x}_i.$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} \left(\exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \right) = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \cdot \mathbf{x}_i.$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} \left(\log(y_i!) \right) = 0 \quad \text{(constante)}.$$

- **Donc, le gradient de la log-vraisemblance est :**

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left[y_i - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \right] \mathbf{x}_i.$$

Solution : MV pour la Régression de Poisson

- Contrairement à la régression linéaire (OLS), il n'y a pas de solution analytique pour :
$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}).$$
- Ceci est dû au fait qu'on ne peut pas résoudre :
$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}.$$
de manière explicite car les dérivées incluent des exponentielles :
$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} [y_i - \mu_i]\, \mathbf{x}_i, \quad \text{avec } \mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}).$$
- $\Rightarrow$ **Il faut utiliser une méthode numérique d'optimisation.**
- On peut utiliser la **descente de gradient** pour trouver $\hat{\boldsymbol{\beta}}$.

Solution : MV pour la Régression de Poisson

- On maximise $\ell(\boldsymbol{\beta})$ en ajustant $\boldsymbol{\beta}$ dans la direction du **gradient** (vecteur score) :

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \alpha \cdot \nabla \ell(\boldsymbol{\beta}^{(t)}).$$

- Le gradient de la log-vraisemblance est donné par:

$$\nabla \ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - \exp(\mathbf{x}_i^\top \boldsymbol{\beta})\right) \mathbf{x}_i = \mathbf{X}^\top (\mathbf{y} - \boldsymbol{\mu}).$$

- Où $\boldsymbol{\mu} = \exp(\mathbf{X}\boldsymbol{\beta})$ est le vecteur des espérances.
- On met à jour $\boldsymbol{\beta}$ jusqu'à convergence.

Solution : MV pour la Régression de Poisson

- **Pourquoi recourir à la descente de gradient?**
- **Pas de solution analytique :** les équations ne se simplifient pas comme pour OLS.
- **Fonction objectif concave :** (la log-vraisemblance est concave pour la régression de Poisson) $\Rightarrow$ toute méthode de gradient converge vers le maximum global.
- **Méthode simple et générale :** fonctionne pour d'autres GLMs (logistique, gamma, etc).
- D'autres alternatives peuvent être efficaces :
  - Méthode de Newton-Raphson
  - Fisher scoring
  - IRLS (Iteratively Reweighted Least Squares)

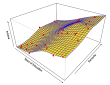
Exercice : Implémentation de la Régression de Poisson en Python sur le Jeu de Donnée *Carseats*

Exercice : Implémentation de la Régression de Poisson

- Dans cet exercice, vous implémenterez manuellement les composantes essentielles d'un modèle de régression de Poisson pour prédire les ventes de sièges pour enfants à partir de variables socio-économiques et commerciales.
- Le jeu de données `Carseats`, issu du manuel *An Introduction to Statistical Learning* (ISLP), contient les caractéristiques de 400 magasins vendant des sièges pour enfants. Chaque ligne correspond à un magasin.
- Le jeu de données vous est fourni sur Moodle. Vous pouvez aussi le télécharger en utilisant la librairie ISLP (`!pip install ISLP`).

Exercice : Implémentation de la Régression de Poisson

- Les variables incluent :
  - `Sales` : Ventes de sièges auto pour enfants (en **milliers d'unités**).
  - `CompPrice`, `Price` : Prix chez le concurrent et prix du magasin.
  - `Income`, `Population` : Revenu moyen, population locale.
  - `Advertising`, `Education`, `Age` : Informations socio-démographiques.
  - `ShelveLoc` : Emplacement du produit (`Good`, `Medium`, `Bad`).
  - `Urban`, `US` : Zone urbaine ? Localisé aux États-Unis ?



Install instructions
Datasets used in ISLP
Auto Data
Bike sharing data
Boston Data
Brain Cancer Data
Caravan
Sales of Child Car Seats
U.S. News and World Report's College Data
Credit Card Balance Data
Credit Card Default Data
Fund Manager Data
Baseball Data
Khan Gene Data
NCI 60 Data
New York Stock Exchange Data
Orange Juice Data

Sales of Child Car Seats

A simulated data set containing sales of child car seats at 400 different stores.

- `Sales` : Unit sales (in thousands) at each location
- `CompPrice` : Price charged by competitor at each location
- `Income` : Community income level (in thousands of dollars)
- `Advertising` : Local advertising budget for company at each location (in thousands of dollars)
- `Population` : Population size in region (in thousands)
- `Price` : Price company charges for car seats at each site
- `ShelveLoc` : A factor with levels Bad, Good and Medium indicating the quality of the shelving location for the car seats at each site
- `Age` : Average age of the local population
- `Education` : Education level at each location
- `Urban` : A factor with levels No and Yes to indicate whether the store is in an urban or rural location
- `US` : A factor with levels No and Yes to indicate whether the store is in the US or not

```
from ISLP import load_data
Carseats = load_data('Carseats')
Carseats.columns
```

```
Index(['Sales', 'CompPrice', 'Income', 'Advertising', 'Population', 'Price',
       'ShelveLoc', 'Age', 'Education', 'Urban', 'US'],
      dtype='object')
```

```
Carseats.shape
```

Exercice : Implémentation de la Régression de Poisson

- Vous devez compléter du code Python dans lequel :
  - Vous implémenterez la log-vraisemblance pour la régression de Poisson.
  - Vous implémenterez le gradient de cette log-vraisemblance.
  - Vous implémenterez la fonction d'entraînement `poisson_regression` qui met à jour les coefficients $\boldsymbol{\beta}$ par descente de gradient.
- La variable cible `Sales` a été arrondie à l'entier le plus proche.
- Les variables catégorielles ont été encodées avec des indicatrices.
- Les données sont divisées en train (80%) et validation (20%).
- La standardisation est faite après le split en utilisant uniquement le train.
- Vous utiliserez la descente de gradient pour maximiser la log-vraisemblance.
- Le clipping est appliqué à $\eta = \mathbf{X}\boldsymbol{\beta}$ pour éviter les overflows numériques.
- Un graphique montrera l'évolution de la log-vraisemblance pour les ensembles d'entraînement et de validation.

Exercice : Implémentation de la Régression de Poisson

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# --- Charger les données Carseats ---
df = pd.read_csv("Carseats_raw.csv")

# --- Transformation de la variable cible ---
df['Sales'] = df['Sales'].round().astype(int)

# --- Encodage des variables catégorielles ---
df_encoded = pd.get_dummies(df, drop_first=False)

# --- Variables explicatives et cible ---
X_vars = [col for col in df_encoded.columns if col != 'Sales']
X_raw = df_encoded[X_vars].values
y = df_encoded['Sales'].values

# --- Division train / validation AVANT standardisation ---
X_raw_train, X_raw_val, y_train, y_val = train_test_split(X_raw, y, test_size=0.2, random_state=8302)

# --- Standardisation basée uniquement sur le train ---
scaler = StandardScaler()
X_train_std = scaler.fit_transform(X_raw_train)
X_val_std = scaler.transform(X_raw_val)

# --- Ajout de l'intercept ---
X_train = np.column_stack((np.ones(X_train_std.shape[0]), X_train_std))
X_val = np.column_stack((np.ones(X_val_std.shape[0]), X_val_std))
variables = ['Intercept'] + X_vars
```

Exercice : Implémentation de la Régression de Poisson

```python
# --- Fonctions log-vraisemblance et gradient ---
def log_likelihood(beta, X, y):
    eta = _______
    eta = np.clip(eta, a_max=30, a_min=None)
    mu = _______
    return _______

def gradient(beta, X, y):
    eta = _______
    eta = np.clip(eta, a_max=30, a_min=None)
    mu = _______
    return _______

# --- Descente de gradient ---
def poisson_regression(X, y, lr=1e-5, max_iter=1000, tol=1e-6, X_val=None, y_val=None):
    beta = np.zeros(X.shape[1])
    ll_history_train = []
    ll_history_val = []

    for iteration in range(max_iter):
        grad = gradient(_______)
        beta = _______

        ll_train = log_likelihood(beta, X, y)
        ll_history_train.append(ll_train)

        if X_val is not None and y_val is not None:
            ll_val = log_likelihood(beta, _______, _______)
            ll_history_val.append(ll_val)

        if iteration > 0 and abs(ll_history_train[-1] - ll_history_train[-2]) < tol:
            print(f"Convergence atteinte en {iteration} itérations.")
            break

    return beta, ll_history_train, ll_history_val
```

Exercice : Implémentation de la Régression de Poisson

```python
# --- Entraînement ---
beta_hat, history_train, history_val = poisson_regression(X_train, y_train, X_val=_______, y_val=_______, lr=1e-5)

# --- Affichage des coefficients estimés ---
print("Coefficients estimés :")
for var, coef in zip(variables, beta_hat):
    print(f"  {var} : {coef:.4f}")

# --- Visualisation 1 : Log-vraisemblance normalisée ---
plt.figure(figsize=(8, 5))
plt.plot(np.array(history_train)/len(y_train), label="Train", color="green")
plt.plot(np.array(history_val)/len(y_val), label="Validation", color="orange")
plt.xlabel("Itérations")
plt.ylabel("Log-vraisemblance normalisée")
plt.title("Évolution de la log-vraisemblance")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

# --- Visualisation 2 : Prédictions vs Observations ---
mu_pred = np.exp(np.clip(X_val @ beta_hat, a_max=30, a_min=None))

plt.figure(figsize=(8, 5))
plt.scatter(mu_pred, y_val, alpha=0.6, color='royalblue')
plt.plot([min(mu_pred), max(mu_pred)], [min(mu_pred), max(mu_pred)], 'r--')
plt.xlabel("Valeurs prédites (mu)")
plt.ylabel("Valeurs observées (Sales)")
plt.title("Régression de Poisson | Prédictions vs Observations")
plt.grid(True)
plt.tight_layout()
plt.show()
```

Solution: Implémentation de la Régression de Poisson en Python sur le Jeu de Donnée *Carseats*

Solution : Implémentation de la Régression de Poisson

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# --- Charger les données Carseats ---
df = pd.read_csv("Carseats_raw.csv")

# --- Transformation de la variable cible ---
df['Sales'] = df['Sales'].round().astype(int)

# --- Encodage des variables catégorielles ---
df_encoded = pd.get_dummies(df, drop_first=False)

# --- Variables explicatives et cible ---
X_vars = [col for col in df_encoded.columns if col != 'Sales']
X_raw = df_encoded[X_vars].values
y = df_encoded['Sales'].values

# --- Division train / validation AVANT standardisation ---
X_raw_train, X_raw_val, y_train, y_val = train_test_split(X_raw, y, test_size=0.2, random_state=8302)

# --- Standardisation basée uniquement sur le train ---
scaler = StandardScaler()
X_train_std = scaler.fit_transform(X_raw_train)
X_val_std = scaler.transform(X_raw_val)

# --- Ajout de l'intercept ---
X_train = np.column_stack((np.ones(X_train_std.shape[0]), X_train_std))
X_val = np.column_stack((np.ones(X_val_std.shape[0]), X_val_std))
variables = ['Intercept'] + X_vars
```

Solution : Implémentation de la Régression de Poisson

```python
# --- Fonctions log-vraisemblance et gradient ---
def log_likelihood(beta, X, y):
    eta = X @ beta
    eta = np.clip(eta, a_max=30, a_min=None)
    mu = np.exp(eta)
    return np.sum(y * eta - mu)

def gradient(beta, X, y):
    eta = X @ beta
    eta = np.clip(eta, a_max=30, a_min=None)
    mu = np.exp(eta)
    return X.T @ (y - mu)

# --- Descente de gradient ---
def poisson_regression(X, y, lr=1e-5, max_iter=1000, tol=1e-6, X_val=None, y_val=None):
    beta = np.zeros(X.shape[1])
    ll_history_train = []
    ll_history_val = []

    for iteration in range(max_iter):
        grad = gradient(beta, X, y)
        beta = beta + lr * grad

        ll_train = log_likelihood(beta, X, y)
        ll_history_train.append(ll_train)

        if X_val is not None and y_val is not None:
            ll_val = log_likelihood(beta, X_val, y_val)
            ll_history_val.append(ll_val)

        if iteration > 0 and abs(ll_history_train[-1] - ll_history_train[-2]) < tol:
            print(f"Convergence atteinte en {iteration} itérations.")
            break

    return beta, ll_history_train, ll_history_val
```

Solution : Implémentation de la Régression de Poisson

```python
# --- Entraînement ---
beta_hat, history_train, history_val = poisson_regression(X_train, y_train, X_val=X_val, y_val=y_val, lr=1e-5)

# --- Affichage des coefficients estimés ---
print("Coefficients estimés :")
for var, coef in zip(variables, beta_hat):
    print(f"  {var} : {coef:.4f}")

# --- Visualisation 1 : Log-vraisemblance normalisée ---
plt.figure(figsize=(8, 5))
plt.plot(np.array(history_train)/len(y_train), label="Train", color="green")
plt.plot(np.array(history_val)/len(y_val), label="Validation", color="orange")
plt.xlabel("Itérations")
plt.ylabel("Log-vraisemblance normalisée")
plt.title("Évolution de la log-vraisemblance")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

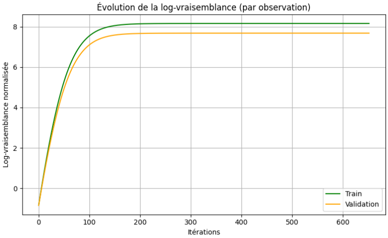
Solution : Implémentation de la Régression de Poisson

```python
# --- Visualisation 2 : Prédictions vs Observations ---
mu_pred = np.exp(np.clip(X_val @ beta_hat, a_max=30, a_min=None))
plt.figure(figsize=(8, 5))
plt.scatter(mu_pred, y_val, alpha=0.6, color='royalblue')
plt.plot([min(mu_pred), max(mu_pred)], [min(mu_pred), max(mu_pred)], 'r--')
plt.xlabel("Valeurs prédites (mu)")
plt.ylabel("Valeurs observées (Sales)")
plt.title("Régression de Poisson | Prédictions vs Observations")
plt.grid(True)
plt.tight_layout()
plt.show()
```

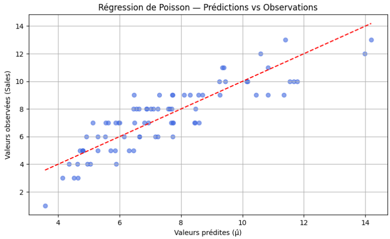
```
Résultat Affiché :
Convergence atteinte en 652 itérations.
Coefficients estimés :
  Intercept : 1.9555
  CompPrice : 0.1939
  Income : 0.0604
  Advertising : 0.0989
  Population : 0.0143
  Price : -0.2953
  Age : -0.1102
  Education : -0.0060
  ShelveLoc_Bad : -0.1294
  ShelveLoc_Good : 0.1335
  ShelveLoc_Medium : 0.0021
  Urban_No : -0.0053
  Urban_Yes : 0.0053
  US_No : 0.0034
  US_Yes : -0.0034
```

Solution : Implémentation de la Régression de Poisson

- La descente de gradient a convergé en 652 itérations, indiquant un bon choix du taux d'apprentissage.
- Les courbes de log-vraisemblance montrent une convergence rapide et l'absence de surapprentissage.
- Les prédictions sur l'ensemble de validation suivent globalement les observations, indiquant une bonne capacité de généralisation.



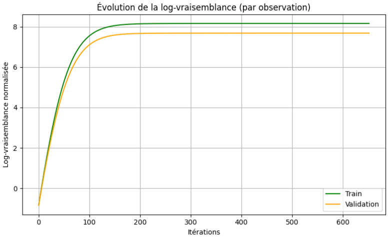
Évolution de la log-vraisemblance



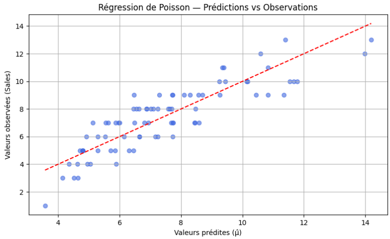
Prédictions vs Observations

Solution : Implémentation de la Régression de Poisson

- Les variables `Price`, `Advertising`, et `ShelveLoc` ont un effet marqué sur les ventes.
- Le coefficient négatif de `Price` est attendu : un prix plus élevé entraîne une baisse des ventes.
- La variable `ShelveLoc_Good` a un effet positif important, illustrant l'impact du placement en rayon.



Évolution de la log-vraisemblance



Prédictions vs Observations

Table des Matières

Regression Logistique

Régression Logistique

Une autre instance des Modèles Linéaires Généralisés (GLMs)

Pourquoi la Régression Logistique ?

- Jusqu'ici, nous avons considéré des variables de sortie $Y \in \mathbb{N}$ (Poisson) ou $Y \in \mathbb{R}$ (Normale).
- **Et si la variable cible est binaire ?**
  Exemple : prédire si un individu est malade ou non, si une transaction est frauduleuse ou non.

$$Y \in \{0, 1\}$$

- **Problème :** La régression linéaire n'est pas adaptée.
  Elle peut produire des prédictions $\hat{y} \notin [0, 1]$, ce qui est incohérent pour une probabilité.
- **Solution :** Utiliser une fonction de lien adaptée :

$$\mu_i = \mathbb{E}[Y_i] = \text{probabilité de succès} \Rightarrow \mu_i \in [0, 1]$$

- C'est ici qu'intervient la **Régression Logistique**, avec fonction de lien logit :

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$$

Exercice : Montrons que la loi de Bernoulli $Y \sim \mathcal{B}(p)$ appartient à la famille exponentielle.

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q1. Quelle est la fonction de masse de $Y \sim \mathcal{B}(p)$?

A. $f_Y(y) = \exp(p^y(1-p)^{1-y})$

B. $f_Y(y) = p^y(1-p)^{1-y}$

C. $f_Y(y) = \frac{1}{p^y(1-p)^{1-y}}$

D. $f_Y(y) = y(1-p) + (1-y)p$

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q1. Quelle est la fonction de masse de $Y \sim \mathcal{B}(p)$?
Réponse correcte : B
La fonction de masse de probabilité d'une variable de Bernoulli est bien :

$$f_Y(y) = p^y(1-p)^{1-y}, \quad y \in \{0, 1\}$$

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q2. Mise sous forme exponentielle : lesquelles sont correctes ?

A. $\exp\left(y\log(p) + (1-y)\log(1-p)\right)$

B. $\exp\left(y\log\left(\frac{p}{1-p}\right) + \log(1-p)\right)$

C. $\exp\left(y\log(p) - y\log(1-p)\right)$

D. $\exp\left(y-p\right)$

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q2. Mise sous forme exponentielle : lesquelles sont correctes ?
Réponses correctes : A et B

$$f_Y(y) = \exp\left(y \log\left(\frac{p}{1-p}\right) + \log(1-p)\right)$$

On reconnaît bien la forme exponentielle.

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q3. Identifier les composantes de la forme exponentielle

A. $\theta = \log\left(\frac{p}{1-p}\right), \quad b(\theta) = \log(1+e^{\theta}), \quad a(\phi) = 1$

B. $\theta = p, \quad b(\theta) = p, \quad a(\phi) = 1$

C. $\theta = \frac{1}{p}, \quad b(\theta) = \theta^2, \quad a(\phi) = \theta$

D. $\theta = \log(p), \quad b(\theta) = \log(1+e^{p}), \quad a(\phi) = 1$

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q3. Identifier les composantes de la forme exponentielle
Réponse correcte : A

- Paramètre canonique : $\theta = \log\left(\frac{p}{1-p}\right)$
- Fonction log-partition : $b(\theta) = \log(1 + e^{\theta})$
- Dispersion : $a(\phi) = 1$ (car Bernoulli à variance fixée)

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q4. Quels sont les moments de $Y \sim \mathcal{B}(p)$?

A. $\mathbb{E}[Y] = b'(\theta), \quad \mathrm{Var}(Y) = b''(\theta)$

B. $\mathbb{E}[Y] = p, \quad \mathrm{Var}(Y) = p(1-p)$

C. $\mathbb{E}[Y] = \frac{e^\theta}{1+e^\theta}, \quad \mathrm{Var}(Y) = \frac{e^\theta}{(1+e^\theta)^2}$

D. Toutes les réponses ci-dessus

Exercice : La Loi de Bernoulli $\in$ Famille Exponentielle ?

Q4. Quels sont les moments de $Y \sim \mathcal{B}(p)$?
Réponse correcte : D
Toutes les réponses sont cohérentes selon l'expression $b(\theta) = \log(1 + e^{\theta})$.

$$b'(\theta) = \frac{e^{\theta}}{1 + e^{\theta}}, \quad b''(\theta) = \frac{e^{\theta}}{(1 + e^{\theta})^2}$$

Solution de l'Exercice : Montrons que la loi de Bernoulli $Y \sim \mathcal{B}(\mu)$ appartient à la famille exponentielle.

Solution : La Loi de Bernoulli $\in$ Famille Exponentielle ?

- Montrons que la loi de Bernoulli $Y \sim \mathcal{B}(\mu)$ **appartient à la famille exponentielle :**

$$f_Y(y;\theta) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

- **Étape 1 : Écrire la densité de la loi de Bernoulli**

$$f_Y(y) = p^y(1-p)^{1-y}, \quad y \in \{0,1\},\ p \in (0,1)$$

- **Étape 2 : Mise sous forme exponentielle, pas à pas**

$$\begin{aligned}
f_Y(y) &= \exp\left(\log(p^y(1-p)^{1-y})\right) \\
&= \exp\left(y\log(p) + (1-y)\log(1-p)\right) \\
&= \exp\left(y\log(p) + \log(1-p) - y\log(1-p)\right) \\
&= \exp\left(y\log\left(\frac{p}{1-p}\right) + \log(1-p)\right)
\end{aligned}$$

Solution : La Loi de Bernoulli $\in$ Famille Exponentielle ?

- **Étape 3 : Identification des termes**
  On reconnaît :

$$f_Y(y) = \exp\left(\underbrace{y \log\left(\frac{p}{1-p}\right)}_{y\theta} + \underbrace{\log(1-p)}_{-b(\theta)} \right) = \exp\left(y\theta - b(\theta) + c(y)\right)$$

- **Identification :**
  - $\theta = \log\left(\frac{p}{1-p}\right)$ (paramètre naturel)
  - $b(\theta) = \log(1 + e^{\theta})$
  - $a(\phi) = 1$ (pas de paramètre de dispersion)
  - $c(y) = 0$

- **Moments :**

$$\mathbb{E}[Y] = b'(\theta) = \frac{e^{\theta}}{1+e^{\theta}} = p, \quad \mathrm{Var}(Y) = a(\phi) \cdot b''(\theta) = \frac{e^{\theta}}{(1+e^{\theta})^2} = p(1-p)$$

Solution : La Loi de Bernoulli $\in$ Famille Exponentielle ?

- On sait que :

$$\mathbb{E}[Y] = b'(\theta) = \frac{e^\theta}{1+e^\theta} = \mu$$

- On inverse cette relation pour obtenir θ en fonction de μ :

$$\mu = \frac{e^\theta}{1+e^\theta} \Rightarrow 1-\mu = \frac{1}{1+e^\theta} \Rightarrow \frac{\mu}{1-\mu} = e^\theta \Rightarrow \theta = \log\left(\frac{\mu}{1-\mu}\right)$$

- **Conclusion :**

$$\boxed{g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)}$$ (fonction de lien canonique)

- Donc le prédicteur linéaire s'écrit :

$$\eta = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad \mu = \frac{e^\eta}{1+e^\eta}$$

Conclusion : Régression Logistique = GLM

Composante du GLM	Régression Logistique
Distribution de Y	$Y \sim \mathcal{B}(\mu)$
Paramètre naturel θ	$\theta = \log\left(\frac{\mu}{1-\mu}\right)$
Paramètre de dispersion ϕ	$\phi = 1$
Fonction de lien $g(\mu)$	$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
Prédicteur linéaire η	$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$
Lien entre μ_i et η_i	$\mu_i = \frac{1}{1+e^{-\eta_i}}$
Méthode d'estimation	Maximum de vraisemblance

Conclusion : La régression logistique est un GLM avec loi de Bernoulli, lien logit.

Table des Matières

Annexe

Dérivation des Moments de la Famille Exponentielle

Dérivation des Moments de la Famille Exponentielle

- Soit la fonction de densité de la famille exponenitelle écrite sous la forme :
$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right).$$
On souhaite démontrer tout d'abord que $\mathbb{E}[Y] = b'(\theta)$.
- **Étape 1 : Calcul de la Log-Vraisemblance**
$$\ell(\theta,\phi \mid y) = \log\left[f_Y(y;\theta,\phi)\right] = \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi).$$
- **Étape 2 : Calcul de la fonction Score** $\frac{\partial \ell}{\partial \theta}$
$$\frac{\partial \ell}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)}.$$
- **Étape 3 — Calcul de l'Espérance du Score** $\mathbb{E}\left[\frac{\partial \ell}{\partial \theta}\right]$
$$\mathbb{E}\left[\frac{\partial \ell}{\partial \theta}\right] = \mathbb{E}\left[\frac{y - b'(\theta)}{a(\phi)}\right] = \frac{\mathbb{E}[Y] - b'(\theta)}{a(\phi)}.$$

Dérivation des Moments de la Famille Exponentielle

- L'espérance du Score est donnée par :

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial \ell}{\partial \theta}\right] &= \mathbb{E}\left[\frac{\log\left[f_Y(y;\theta,\phi)\right]}{\partial \theta}\right] \\
&= \int \frac{1}{f_Y(y;\theta,\phi)} \cdot \frac{\partial f_Y(y;\theta,\phi)}{\partial \theta} \cdot f_Y(y;\theta,\phi)\, dy \\
&= \int \frac{f_Y(y;\theta,\phi)}{f_Y(y;\theta,\phi)} \cdot \frac{\partial f_Y(y;\theta,\phi)}{\partial \theta} \cdot\, dy \\
&= \int \frac{\partial f_Y(y;\theta,\phi)}{\partial \theta}\, dy \\
&= \frac{\partial}{\partial \theta} \int f(y;\theta,\phi) dy \\
&= \frac{\partial}{\partial \theta}(1) = 0
\end{aligned}
$$

- $\mathbb{E}\left[\frac{\partial \ell}{\partial \theta}\right] = 0 \Rightarrow \frac{\mathbb{E}[Y]-b'(\theta)}{a(\phi)} = 0 \Rightarrow \mathbb{E}[Y] = b'(\theta)$ (CQFD).

Dérivation des Moments de la Famille Exponentielle

- On souhaite maintenant démontrer que $\text{Var}(Y) = b''(\theta) \cdot a(\phi)$ en exploitant la dérivée seconde de la log-vraisemblance.
- Fonction log-vraisemblance :

$$\ell(\theta, \phi \mid y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

- Fonction score :

$$\frac{\partial \ell}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)}$$

- Dérivée seconde :

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)}, \quad \textbf{Résultat (I).}$$

- On a besoin de démontrer l'identité de **l'Information de Fisher** :

$$\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]$$

Dérivation des Moments de la Famille Exponentielle

- On rappelle que la fonction log-vraisemblance est :

$$\ell(\theta, \phi \mid y) = \log\left[f_Y(y; \theta, \phi)\right]$$

- On cherche à calculer :

$$\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log\left[f_Y(y; \theta, \phi)\right]\right]$$

- On applique la dérivée du logarithme d'une fonction :

$$\frac{\partial}{\partial \theta} \log(f_Y) = \frac{1}{f_Y} \cdot \frac{\partial f_Y}{\partial \theta}$$

- Donc, la dérivée seconde s'écrit :

$$\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \frac{\partial^2}{\partial \theta^2} \log(f_Y) = \frac{\partial}{\partial \theta}\left(\frac{1}{f_Y} \cdot \frac{\partial f_Y}{\partial \theta}\right)$$

Dérivation des Moments de la Famille Exponentielle

- On applique la règle du quotient et de la chaîne :

$$\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \frac{\partial}{\partial \theta}\left(\frac{1}{f_Y} \cdot \frac{\partial f_Y}{\partial \theta}\right) = \frac{\partial}{\partial \theta}\left(f_Y^{-1} \cdot \frac{\partial f_Y}{\partial \theta}\right)$$
$$= \left(-\frac{1}{f_Y^2} \cdot \frac{\partial f_Y}{\partial \theta}\right) \cdot \frac{\partial f_Y}{\partial \theta} + \frac{1}{f_Y} \cdot \frac{\partial^2 f_Y}{\partial \theta^2}$$
$$= -\frac{1}{f_Y^2}\left(\frac{\partial f_Y}{\partial \theta}\right)^2 + \frac{1}{f_Y} \cdot \frac{\partial^2 f_Y}{\partial \theta^2}$$

- Ainsi, on a :

$$\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \mathbb{E}\left[-\frac{1}{f_Y^2}\left(\frac{\partial f_Y}{\partial \theta}\right)^2 + \frac{1}{f_Y} \cdot \frac{\partial^2 f_Y}{\partial \theta^2}\right]$$
$$\Rightarrow \mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = -\mathbb{E}\left[\left(\frac{1}{f_Y} \cdot \frac{\partial f_Y}{\partial \theta}\right)^2\right] + \mathbb{E}\left[\frac{1}{f_Y} \cdot \frac{\partial^2 f_Y}{\partial \theta^2}\right]$$
$$= -\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] + \mathbb{E}\left[\frac{1}{f_Y} \cdot \frac{\partial^2 f_Y}{\partial \theta^2}\right]$$

Dérivation des Moments de la Famille Exponentielle

- On remarque que :

$$\mathbb{E}\left[\frac{1}{f_Y} \cdot \frac{\partial^2 f_Y}{\partial \theta^2}\right] = \int \frac{1}{f_Y} \cdot \frac{\partial^2 f_Y}{\partial \theta^2} \cdot f_Y \, dy = \int \frac{\partial^2 f_Y}{\partial \theta^2} \, dy = \frac{\partial^2}{\partial \theta^2} \int f_Y(y;\theta,\phi) \, dy = \frac{\partial^2}{\partial \theta^2}(1) = 0$$

- Donc :

$$\mathbb{E}\left[\frac{1}{f_Y(y;\theta,\phi)} \cdot \frac{\partial^2 f_Y(y;\theta,\phi)}{\partial \theta^2}\right] = 0$$

- Il en résulte que :

$$\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = -\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right],$$ **Résultat (II).**

Dérivation des Moments de la Famille Exponentielle

- On a :

$$\frac{\partial \ell}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \quad \Rightarrow \quad \left(\frac{\partial \ell}{\partial \theta}\right)^2 = \frac{(y - b'(\theta))^2}{a(\phi)^2}$$

- Alors :

$$\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = \frac{\mathbb{E}\left[(Y-\mu)^2\right]}{a(\phi)^2} = \frac{\mathrm{Var}(Y)}{a(\phi)^2}$$

- D'autre part, d'après **Résultat (I)** :

$$-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \frac{b''(\theta)}{a(\phi)}.$$

- En égalant les deux expressions dans **Résultat (I)** et **Résultat (II)**:

$$\frac{b''(\theta)}{a(\phi)} = \frac{\mathrm{Var}(Y)}{a(\phi)^2} \Rightarrow \mathrm{Var}(Y) = b''(\theta) \cdot a(\phi), \quad \text{(CQFD)}.$$

Fonction de Score et Information de Fisher

Fonction de Score et Information de Fisher

- Toute variable aléatoire Y dont la loi appartient à la famille exponentielle a une densité de probabilité de la forme :

$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

- Dans ce cadre, on peut analyser :
  - **la fonction de score**, qui mesure la sensibilité de la log-vraisemblance par rapport à θ,
  - **l'information observée**, qui donne la courbure locale de la log-vraisemblance,
  - **l'information de Fisher**, qui mesure la précision espérée de l'estimateur.

Fonction de Score et Information de Fisher

- **Définition formelle :** Soit $Y \sim f_Y(y;\theta)$, une variable aléatoire dont la loi dépend d'un paramètre $\theta \in \mathbb{R}$. La log-vraisemblance est donnée par:

$$\boxed{\ell(\theta; y) = \log\left[f_Y(y;\theta)\right]}$$

  La **fonction de score** $U(\theta)$ est la dérivée de la log-vraisemblance par rapport à θ :

$$\boxed{U(\theta) := \frac{\partial}{\partial\theta}\ell(\theta; y) = \frac{\partial}{\partial\theta}\log\left[f_Y(y;\theta)\right]}$$

- $U(\theta)$ est une variable aléatoire (car dépend de Y) :
  - Elle indique la direction dans laquelle modifier θ pour augmenter la vraisemblance.

Fonction de Score et Information de Fisher

- $U(\theta) = \frac{\partial \ell}{\partial \theta}$ agit comme une **boussole** qui guide l'optimisation de θ.
- **Intuition** : si on observe une donnée y, on se demande *Pour quelle valeur de θ cette donnée est-elle la plus probable ?*
  - Si $\frac{\partial \ell}{\partial \theta} > 0$: on peut augmenter θ pour améliorer la vraisemblance.
  - Si $\frac{\partial \ell}{\partial \theta} < 0$: on peut diminuer θ.
  - Si $\frac{\partial \ell}{\partial \theta} = 0$: on est sur un **point critique** (potentiellement maximum).
- **Résultat fondamental :** On a démontré dans **le calcul de la moyenne pour une densité faisant partie de la famille exponentielle que**

$$\mathbb{E}[U(\theta)] = \mathbb{E}\left[\frac{\partial \ell}{\partial \theta}\right] = 0.$$

- **Ceci est vrai pour n'importe quelle densité $f_Y(y;\theta)$:**
  - Le maximum de vraisemblance est atteint lorsque la dérivée de la log-vraisemblance est nulle en moyenne.

Fonction de Score et Information de Fisher

- **Définition** : **L'information observée** est l'opposée de la dérivée seconde de la log-vraisemblance :

$$\boxed{J(\theta) := -\frac{\partial^2 \ell(\theta; y)}{\partial \theta^2}}$$

- C'est une mesure locale de la courbure de la log-vraisemblance autour de θ.
- Si la log-vraisemblance est très courbée (pic étroit), alors l'information est grande en quantité et l'estimation est alors plus précise.
- Si elle est plate, alors l'estimation est plus incertaine.

Fonction de Score et Information de Fisher

- **L'information de Fisher** est l'espérance de l'information observée. Elle quantifie cette concavité moyenne :

$$\boxed{I(\theta) := \mathbb{E}_\theta\left[J(\theta)\right] = -\mathbb{E}_\theta\left[\frac{\partial^2 \ell(\theta; Y)}{\partial \theta^2}\right]}$$

- Lorsque l'on peut échanger dérivation et intégration, on a aussi :

$$\boxed{I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial \ell(\theta; Y)}{\partial \theta}\right)^2\right]}$$

- On a démontré dans **le calcul de la variance pour une densité faisant partie de la famille exponentielle que** que les deux définitions sont équivalentes et ceci est vrai pour n'importe quelle densité $f_Y(y; \theta)$.

Fonction de Score et Information de Fisher

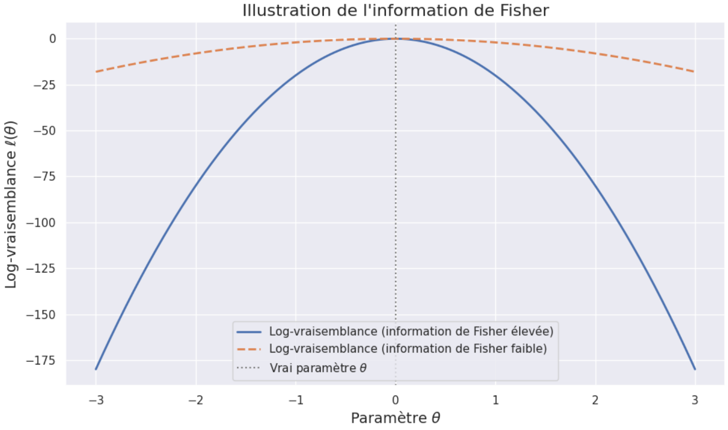
- Étant donné que $\mathbb{E}\left[\frac{\partial \ell}{\partial \theta}\right] = 0$, l'information de Fisher exprime la variance du score $\text{Var}[U(\theta)]$ puisque :

$$
\begin{aligned}
\text{Var}\left(\underbrace{\frac{\partial \ell(\theta;Y)}{\partial \theta}}_{U(\theta)}\right) &= \mathbb{E}_\theta\left[\left(\frac{\partial \ell(\theta;Y)}{\partial \theta}\right)^2 - \left(\underbrace{\mathbb{E}_\theta\left[\frac{\partial \ell(\theta;Y)}{\partial \theta}\right]}_{0}\right)^2\right] \\
&= \mathbb{E}_\theta\left[\left(\frac{\partial \ell(\theta;Y)}{\partial \theta}\right)^2\right] \\
&= \mathbb{E}_\theta\left[J(\theta)\right] \\
&= I(\theta)
\end{aligned}
$$

- **Interprétation intuitive :**
  - $I(\theta)$ mesure la précision moyenne de notre estimation.
  - Plus $I(\theta)$ est grand, plus la log-vraisemblance est "pointue", et plus on estime θ précisément.

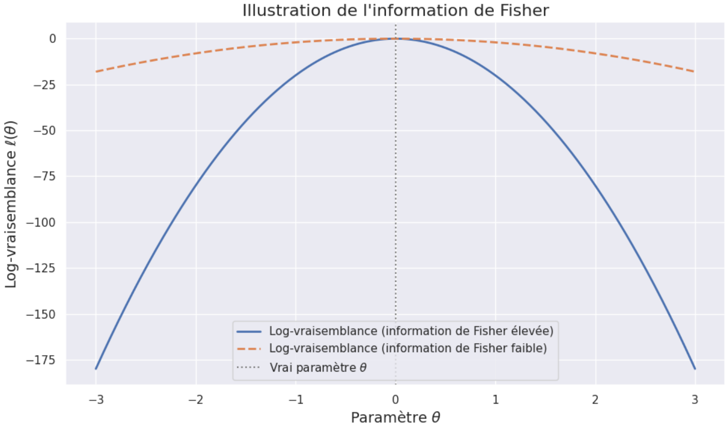
Fonction de Score et Information de Fisher

- **Interprétation intuitive :**
  - $I(\theta)$ mesure la précision moyenne de notre estimation.
  - Plus $I(\theta)$ est grand, plus la log-vraisemblance est "pointue", et plus on estime θ précisément.



Fonction de Score et Information de Fisher

- **Interprétation intuitive :**
  - Exemple : En régression linéaire où $Y \sim \mathcal{N}(\mu, \sigma^2)$, $\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \frac{b''(\theta)}{a(\phi)} = \frac{1}{\sigma^2}$. Plus le bruit est faible, plus l'estimation est précise.



Fonction de Score et Information de Fisher

Terme	**Expression**	**Interprétation**
Fonction de score $U(\theta)$	$\frac{\partial}{\partial\theta} \log f_Y(y;\theta)$	Sensibilité de la log-vraisemblance
Information observée $J(\theta)$	$-\frac{\partial^2}{\partial\theta^2} \log f_Y(y;\theta)$	Courbure locale
Information de Fisher $I(\theta)$	$\mathbb{E}[J(\theta)] = \text{Var}(U(\theta))$	Précision moyenne de l'estimation de θ